2009 Annual Report

Project Title: Engineering metabolic channels as a new approach to microbial chemical manufacture (Young Investigator Program)
ONR Award Number: N00014-06-1-0565
PI: Matthew DeLisa, Cornell University

A. Scientific and Technical Objectives

Our long-term goal is to use tools from molecular biology to engineer multi-enzyme metabolic complexes, mimicking the physical forms ubiquitous in nature. The direct coupling between sequential enzymatic reactions, through either static or dynamic interactions, offers the promise of eliminating these production barriers as it reduces the distance between enzyme active sites and favors sequential reactions over diffusion into the bulk. Therefore, the objective of these studies is to engineer synthetic metabolic complexes by exploiting the assembly mechanisms of natural systems to spatially organize enzymes that participate in sequential reaction steps. We *expect* that by developing a generic set of tools to co-localize metabolic pathways, we will overcome traditional bottlenecks that limit the commercial viability of microbial factories. We have proposed the following specific aims:

(1) To demonstrate efficient multi-protein assembly in bacterial cells. To enable the intentional engineering of metabolic enzymes into functional metabolic complexes, we will explore a variety of novel methods for *in vivo* enzyme assembly.

(2) To assemble functional metabolic complexes. We will utilize the intracellular assembly/cross-linking methodologies from Aim 1 to create synthetic metabolic complexes in bacteria that are capable of efficient metabolic conversions via fermentation of renewable resources. We have chosen as a model system the microbial synthesis of propylene glycol (1,2-propanediol or 1,2-PD).

(3) To enable combinatorial engineering of metabolic complexes via metabolite sensors. We will engineer a collection of intracellular switches that are capable of dynamically responding to intracellular metabolites (e.g., 1,2-PD) over a broad concentration range.

B. Approach

For Aims 1 and 2, our approach is to develop in vivo protein assembly/crosslinking strategies using modern molecular biology techniques. The first approach is direct translational fusion of the enzyme sequences resulting in a single covalently cross-linked fusion protein. The second approach is to employ a unique enzyme known as transglutaminase (TGase) that catalyzes the post-translational modification of substrate proteins by the formation of covalent isopeptide bonds. A third approach is to graft “protein interacting domains” onto each pathway enzyme, thereby creating artificial interaction domains that promote intracellular enzyme assembly. These different assembly/cross-linking techniques will be developed in the context of the three-enzyme pathway that comprises 1,2-PD production in *Escherichia coli*. This entails cloning of the requisite constructs, expression in *E. coli* and examination of the 1,2-PD titers using HPLC/Mass Spec/NMR.

For Aim 3, our approach to develop metabolite sensors using protein-based switches that elicit a measurable activity upon small molecule binding. This is a departure from the original proposal as we have learned from experimentation that RNA-based switches are incapable of detecting small molecules that lack steric bulk. These protein-based switches will be useful for direct monitoring of intracellular 1,2-PD titers in living cells and are expected to open the door to laboratory evolution of our engineered metabolic complexes.

Report Documentation Page

Form Approved
OMB No. 0704-0188

1. REPORT DATE	2. REPORT TYPE	3. DATES COVERED
2009	**N/A**	**-**

4. TITLE AND SUBTITLE	5a. CONTRACT NUMBER
Engineering metabolic channels as a new approach to microbial chemical manufacture	5b. GRANT NUMBER
	5c. PROGRAM ELEMENT NUMBER
6. AUTHOR(S)	5d. PROJECT NUMBER
	5e. TASK NUMBER
	5f. WORK UNIT NUMBER
7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES) **Cornell University**	8. PERFORMING ORGANIZATION REPORT NUMBER
9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)	10. SPONSOR/MONITOR'S ACRONYM(S)
	11. SPONSOR/MONITOR'S REPORT NUMBER(S)

12. DISTRIBUTION/AVAILABILITY STATEMENT
Approved for public release, distribution unlimited

13. SUPPLEMENTARY NOTES
The original document contains color images.

14. ABSTRACT

15. SUBJECT TERMS

16. SECURITY CLASSIFICATION OF:			17. LIMITATION OF ABSTRACT	18. NUMBER OF PAGES	19a. NAME OF RESPONSIBLE PERSON
a. REPORT	b. ABSTRACT	c. THIS PAGE			
unclassified	**unclassified**	**unclassified**	**SAR**	**9**	

Standard Form 298 (Rev. 8-98)
Prescribed by ANSI Std Z39-18

C. Concise Accomplishments

The major accomplishments to date include:

(1) We have demonstrated 1,2-PD production in *E. coli*. Using this pathway as a model system, we have performed a thorough analysis of synthetic enzyme assembly methods mediated by (1) translational fusions; (2) protein interacting domains (PIDs) and (3) eukaryotic signaling scaffolds. A patent application has been disclosed and a manuscript has been submitted.

(2) We have developed a chemical genetic reporter of protein stability that enables intracellular sensing of small compounds (described in progress report for N00014-07-1-0027). The assay was used to identify candidate 1,2,4-butanetriol (D-BT) biosensors. Preliminary efforts to isolate similar 1,2-PD biosensors are underway. The significance of this tool, especially if it can be generalized to multiple different metabolites, is that it provides a fluorescence-based reporter of metabolite levels thereby opening the door to laboratory evolution of our metabolic assemblies.

D. Expanded Accomplishments

Summary. To date, we have engineered *E. coli* for 1,2-PD biosynthesis. Using this pathway as a model system, we have co-localized the 1,2-PD biosynthetic enzymes through both covalent as well as dynamic interactions. In the latter case, we have created modular scaffolds to which any number of pathway enzymes can be assembled. We have demonstrated that each metabolic channel towards 1,2-PD is more efficient than its free enzyme counterpart on a per enzyme basis in the cytoplasm of *E. coli*. Finally, since genetic reporters for most intracellular metabolites are lacking, we have developed a procedure for engineering protein switches that are capable of dynamically responding to a broad concentration range of specific metabolites. We are in the early stages of applying this strategy to create biosensors for 1,2-PD production.

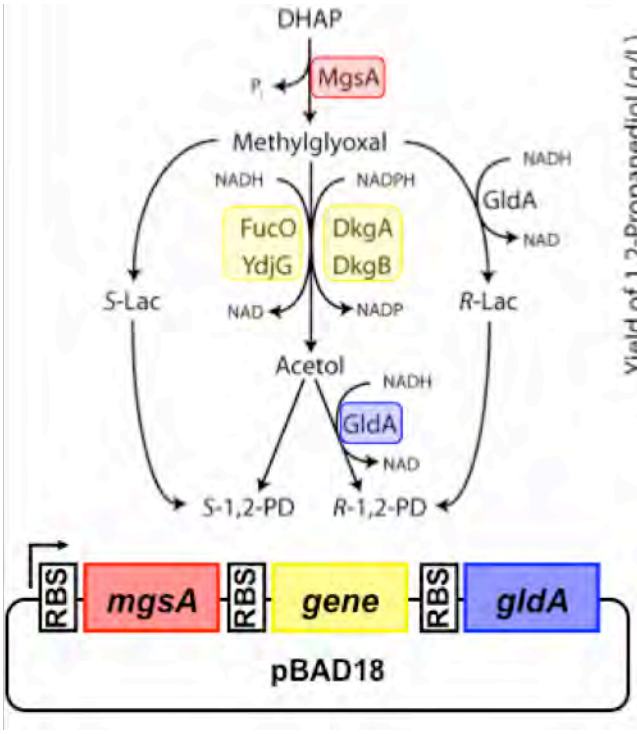

Figure 1. (top) Metabolic pathway for 1,2-PD production. (bottom) Plasmid design for co-expressing 1,2-PD metabolic enzymes in an unfused format.

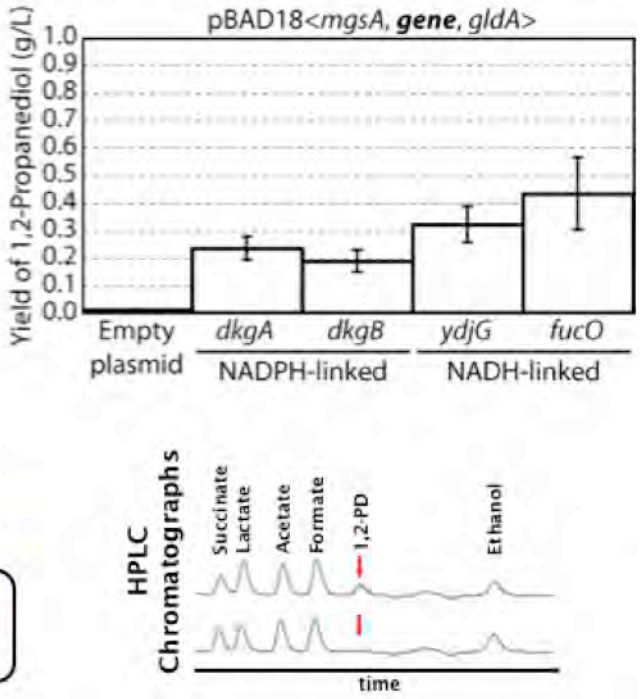

Figure 2. MG reductase enzymes. 1,2-PD production levels for NAD(P)H dependent enzymes coexpressed with *E. coli* MgsA and GldA within pBAD18.

D1. Engineering 1,2-PD biosynthesis in *E. coli.* We have successfully constructed a synthetic metabolic pathway from DHAP to *R*-1,2-PD (Fig. 1, top). Along this path, the first and last enzymatic steps were well-defined, specifically the synthesis of MG by *E. coli* methylgloxal synthase (MgsA) and reduction of acetol by *E. coli* glycerol dehydrogenase (GldA) (Altaras & Cameron, 1999). However, the reduction of the intermediate, MG to acetol, can be performed by a number of NADH or NADPH dependent enzymes, most of which have not been tested *in vivo*. To test for activity on MG *in vivo*, we designed a high copy plasmid expressing MgsA, one of the following NADH (FucO, YdjG) or NADPH dependent enzymes (DkgA, DkgB), and GldA (Fig. 1, bottom). This was achieved in plasmid pBAD18, an arabinose inducible plasmid of pBR322 origin, where the three genes were translated from a polycistronic mRNA, each under control of separate but identical ribosome binding sites. Plasmids were transformed into wild-type *E. coli* strain MC4100, subcultured anaerobically in the presence of glucose, and extracellular levels of 1,2-PD were measured after fermentation by HPLC analysis (Fig. 2). All genetic constructs produced significant levels of 1,2-PD with the NADH dependent enzymes showing higher activity towards MG demonstrated by higher 1,2-PD levels (Fig. 2, top). In moving forward, we chose FucO and DkgA, the highest NADH and NADPH dependent enzymes respectively.

D2. Enzyme compartmentalization towards production of 1,2-PD. In our initial attempts to co-localize metabolic enzymes, we investigated (1) protein fusions and (2) grafting of protein-protein interaction domains (PIDs) onto metabolic enzymes of interest. The latter approach results in the most significant improvement in 1,2-PD titers compared to cells expressing unassembled enzymes (Fig. 3). This is significant as it confirms that enzyme organization is a powerful approach for developing highly efficient metabolic machinery in *E. coli*.

For this mode of assembly, we employed known protein interacting domains, that when fused N- or C-terminally to our pathway enzymes, would bind together and would thus bring the sequential active sites into close proximity (Fig. 3). This strategy offers more design flexibility than simple gene fusion because, in addition to design of the linker region, there are also considerations for the binding associations, homo- or heterodimerization, and their affinity in terms of the K_D. Here our design criteria were domains that: (1) are well expressed in *E. coli*, (2) possess high affinity interactions in the μM to sub-μM range, given that natural proteins can reach 1-10 μM concentrations in *E. coli* (Sceller *et al.*, 2000), (3) are short in length, preferably <50 amino acids, (4) are characterized by highly specific interactions with little cross-reactivity. Based on these standards, we selected 3 sets of interacting domains derived from eukaryotic transcription factors (GCN4 and cJun/cFos) and signaling proteins (e.g., SH3) for further analysis (Table 1).

Table 1. Protein Interacting Domains and their Properties

Name	Type	Association Type	Size (AAs)	K_D (μM)	Refs.
GCN4/GCN4	Leucine Zipper	Homodimerization	48 / 48	0.5	(Dragan et al., 2004)
cJun/cFos	Leucine Zipper	Heterodimerization	41 / 40	0.001-0.11	(Pernelle et al., 1993, Patel et al., 1994, Oyama et al., 2006, Heuer et al., 1996)
SH3/ SH3ligand	Protein Interacting Domains	Heterodimerization	57 / 11,8	0.1, 10	(Dueber et al., 2007)

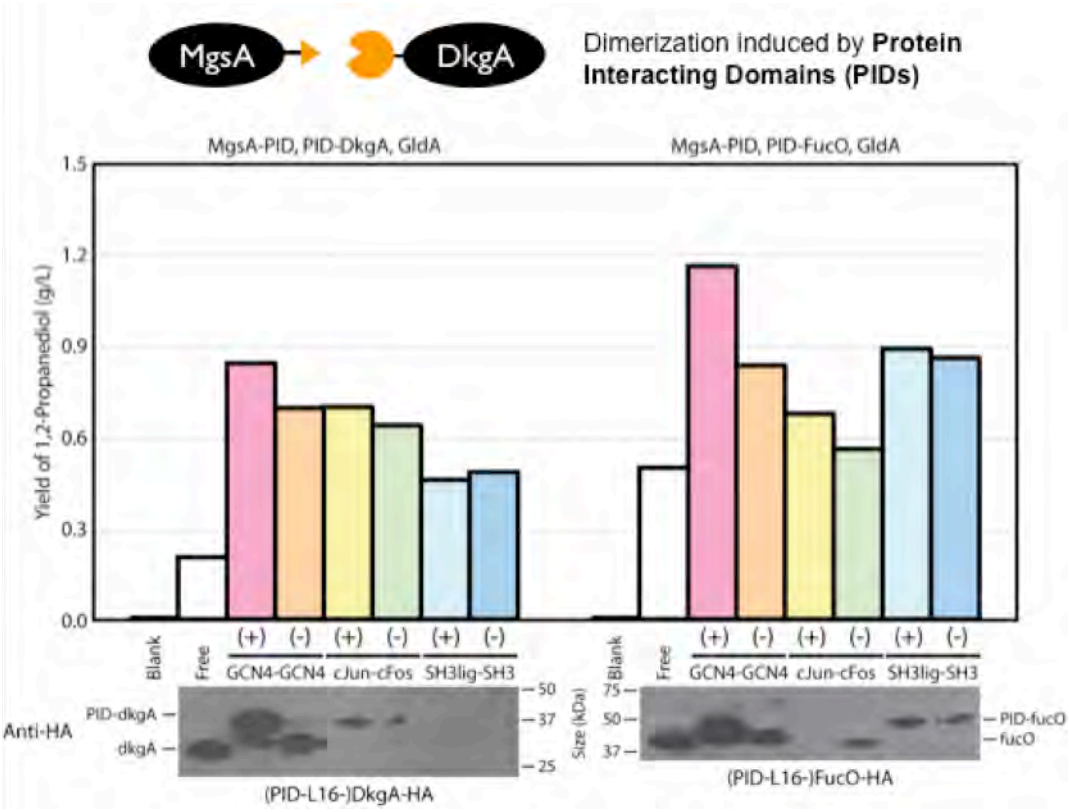

Figure 3. Enzyme co-localization using PIDs. 1,2-PD production using several PID systems, listed in Table 2, as well as the empty plasmid and free enzyme control. PIDs are fused C-terminally to MgsA and N-terminally to DkgA or FucO. GldA was coexpressed. Western blots of soluble fraction against Anti-HA antibody. (+) indicates wild-type PIDs and (-) indicates PIDs with sequences that decreases binding affinity.

These domains were fused to each of the first two pathway enzymes with the L16 linker in a polycistronic mRNA with the third enzyme freely expressed (e.g., MgsA-L16-cJun; cFos-L16-DkgA; GldA). In this manner, we were able to measure the degree of synergistic effect by comparison to a polycistronic free enzyme system. Additionally, the PIDs offer further controls in that in each of these interacting domains, (1) point mutations can be made to reduce binding affinity and (2) cross-reacting species can be combined (e.g., MgsA-SH3; cFos-DkgA). These additional controls should help to elucidate the mechanism of the fold-improvement seen in protein fusions, namely whether the increase was due to enzyme compartmentalization or an increase in protein stability as seen by Western blot. These constructs were analyzed as above for their extracellular 1,2-PD production levels following fermentation (Fig. 3), with each of the interacting domains (+) and their lower affinity counterparts (-). Interestingly for the DkgA or FucO system, the GCN4 leucine zippers were responsible for the largest increase in production, followed by either cJun/cFos or the SH3/SH3ligand pair. For the FucO system, the GCN4 PIDs resulted in the largest increase observed thus far, exceeding our earlier results with the fusion proteins. The Western blot analysis on these samples revealed that the GCN4 fusions significantly increased the stability of enzymes. However, the cJun/cFos and the SH3 fusions had significantly reduced protein levels, despite their relatively high activity. This suggests that enzyme compartmentalization plays a strong role in these cases. Additionally, although the mutations in these last 2 cases reduce the affinity 100-fold, the K_D values may still be below the intracellular enzyme concentrations, which might allow metabolic channeling to be active here. Based on these encouraging results, we constructed a 3-enzyme interacting system (i.e., MgsA-GCN4; GCN4-DkgA/FucO-cJun; cFos-GldA). However, this yielded lower 1,2-PD levels than the

original 2-enzyme systems (results not shown). We believe this is due to the fact that the middle enzymes, DkgA and FucO, are poorly expressed/folded in the presence of both GCN4 and cFos domains at the N- and C-termini, respectively.

D3. Synthetic enzyme scaffolds for co-localizing 1,2-PD pathway enzymes. While useful for systems of 2 enzymes, fusion proteins and interacting domains fall short when trying to assemble entire metabolic pathways comprised of 3 or more enzymes. The problem seems to hinge at the protein instability that results from large fusions at both the N- and C-termini of the native protein sequence. To overcome this, we have created a stable scaffold *in vivo* with several highly specific and modular docking sites to which successive pathway enzymes can be non-covalently attached for enhanced metabolite production (DeLisa & Conrado, 2009). Our strategy is based on a chimeric protein scaffold composed of various eukaryotic protein domains (Table 2) that has been tested both *in vitro* for signaling applications (Dueber et al., 2007) and *in vivo* for metabolite production (DeLisa & Conrado, 2009, Dueber *et al.*, 2009). Pathway enzymes are co-localized on this scaffold by several short ligands, which specifically target each of these docking domains. Specifically, the ligands that correspond to the SH3-PDZ-GBD scaffold were fused N- or C-terminally to different 1,2-PD pathway enzymes. Positioning of

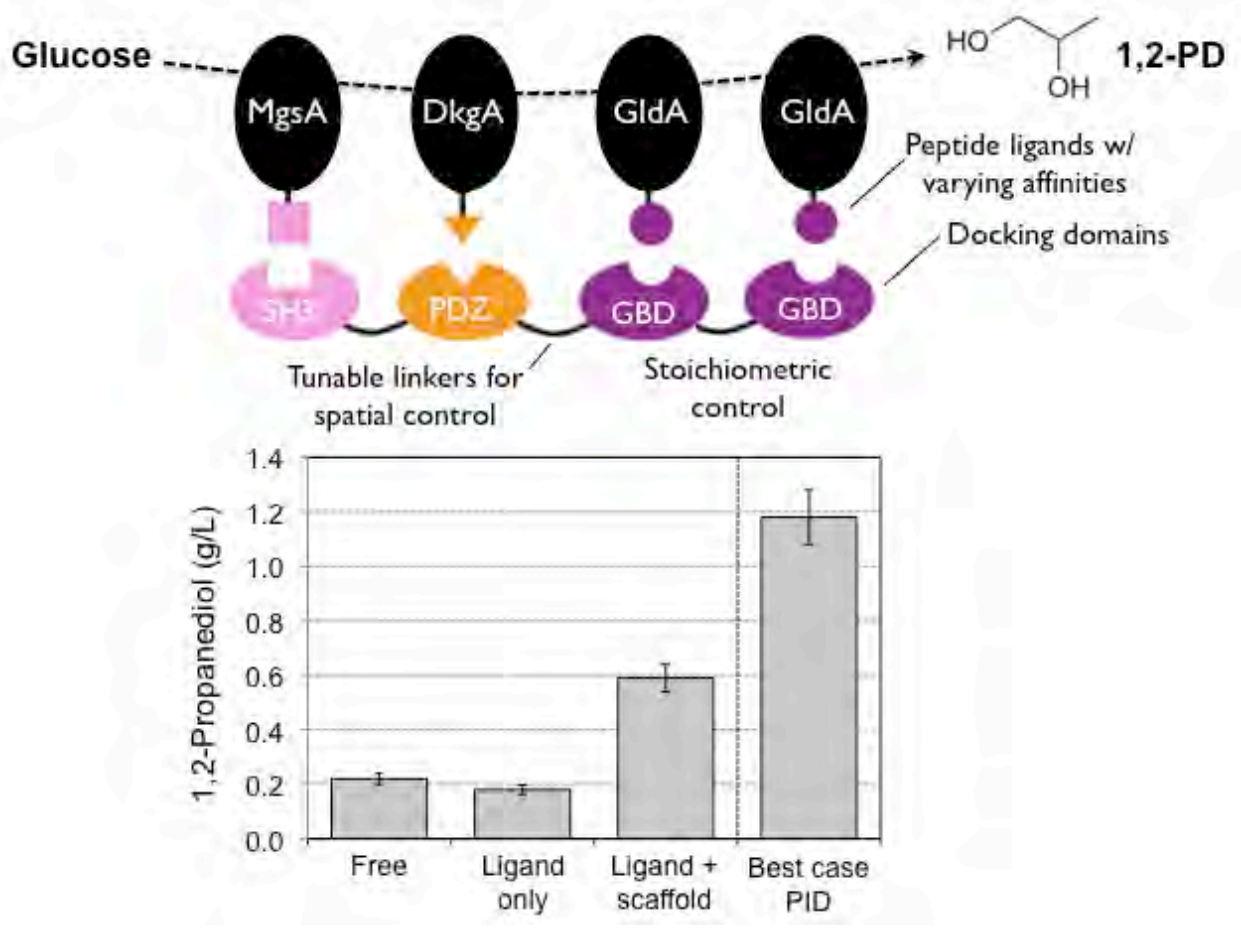

Figure 4. Synthetic protein scaffolds for assembling metabolic enzymes. (top) 1,2-PD production using a protein scaffold derived from eukaryotic signaling proteins and their corresponding small ligands (Table 2) fused to 1,2-PD pathway enzymes. (bottom) 1,2-PD titers from cells expressing unassembled, unmodified enzymes (free), enzyme-ligand constructs without the scaffold (ligand only) and with the scaffold (ligand + scaffold), and the first and second enzymes assembled by GCN4 PIDs (best case PID). Data is the average of 3 replicate experiments where the error is the standard error of the mean.

the ligands with respect to each of the enzymes was based on our earlier studies that identified optimal linker composition/length and permissive orientation (i.e., N- or C-terminus). In the metabolic pathway towards 1,2-PD, MgsA was targeted to the SH3 domain, FucO to the PDZ domain, and GldA to the GBD domain (Fig. 4, schematic). Following co-expression of the scaffold and each of the enzyme-ligand constructs, we measured a nearly 3-fold improvement in 1,2-PD titers relative to the case where the enzymes were expressed separately with no intentional assembly (Fig. 4). It is noteworthy that addition of the small ligands did not affect the activity of the individual enzymes because the enzyme-ligand constructs expressed without the scaffold were nearly as active as the unassembled, unmodified enzymes (Fig. 4). Importantly, the improvement in 1,2-PD conferred by the protein scaffold was only about 50% as effective as our best case using the GCN4 PIDs attached to 2 of the pathway enzymes (Fig. 4). Thus, further optimization of the enzyme-scaffold system is needed. Nonetheless, the development of this scaffold provides a generic framework for co-localization of virtually any metabolic pathway because the required fusion to the sequential enzymes is short, and unlikely to interfere with protein folding and enzyme activity. This will be confirmed by extending the same scaffolding strategy to the production of an unrelated compound, D-BT, under grant N00014-07-1-0027. Note that for the longer pathway towards D-BT (described in progress report for N00014-07-1-0027), the final 3 enzymes will be targeted to the scaffold since the final enzymatic steps suffer from several competing reactions. Thus, YjhG will be targeted to SH3, MdlC to the PDZ domain, and AdhP to the GBD domain by fusing the strong ligands listed in Table 2 to the C-termini of each pathway enzyme by means of the L16 linker.

Table 2. Protein Scaffold Components and their Properties (reproduced from (Dueber et al., 2007)

Docking domain	Source	Ligand (AAs)	K_D (μM)
SH3	Mouse Crk (134-191)	PPPALPPKRRR (11)	0.1
		PPPVPPRR (8)	10
PDZ	Mouse a-syntrophin (77-171)	GVKESLV (7)	8
		GVKQSLL (7)	100
GBD	Rat N-WASP (196-274)	SGIVGALMEVMQKRSKAIH (19)	1

E. Work Plan

E1. New approaches to protein assembly - DNA scaffolds. As an alternative to co-localizing metabolic pathways along a protein scaffold, we will also develop a novel DNA-based scaffold which might prove to be more abundant, stable, and amenable to alteration (Fig. 5). Here a DNA-based scaffold is an attractive alternative as these stable genetic elements (e.g. plasmid DNA) are well defined and easily modified, and further, there are many known DNA binding proteins that can target our pathway enzymes along the DNA platform. Specifically a number of zinc finger domains have been evolved for high specificity towards 49 of the 64 three-base pair (bp) recognition sequences. These have been genetically combined to create multidomain proteins capable of targeting 18 base pair sites that can discriminate single base pair changes with up to 100-fold loss in affinity (Mandell & Barbas, 2006, Segal *et al.*, 1999, Liu *et al.*, 1997). We will employ this technique to create promoterless high-copy plasmids that encode interspaced docking sites, here only 12 base pairs in length. Each zinc finger domain, 22 amino acids in length, targets 3 sequential DNA base pairs; therefore we will fuse 4 zinc finger elements together to create motif capable to specifically targeting our plasmid docking sites, without suffering from competition of genomic DNA or size instability. This technique will enable the co-localization of longer metabolic pathways, e.g. D-BT biosynthesis, and will enable further engineering improvements by controlling the spacing of the sequential enzymes and their geometric arrangements. This will enable multiple repeats of slow enzymes within a co-localized

element, and also several repeats of the entire pathway along a single plasmid DNA *in vivo* to along for toggling the number of channels.

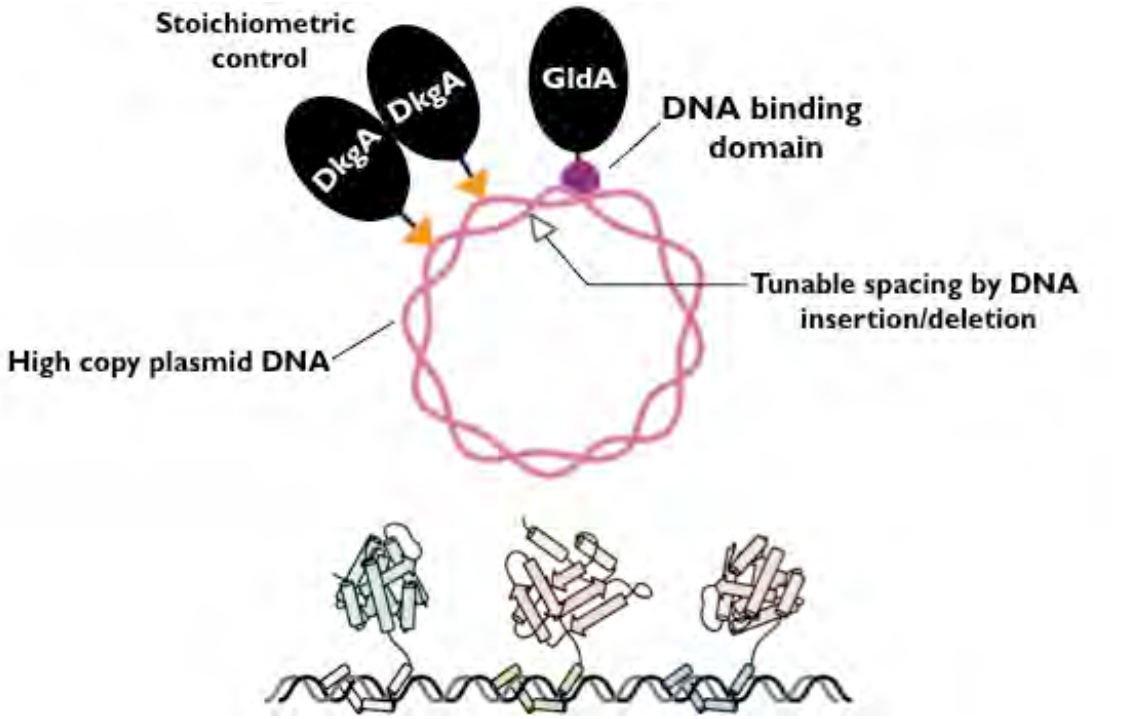

Figure 5. DNA scaffold from a plasmid, showing the protein-binding region, spaced 12 base pair elements. Metabolic enzymes are colocalized along the DNA by zinc finger domains fused C-terminally.

E2. Revisiting which enzymes to assemble. During the course of our 1,2-PD studies, we noticed that cells expressing pathway enzymes (unassembled or assembled) grew slowly. We speculated this was because DHAP accumulation in these cells was limiting metabolism and thus slowing cell growth. In support of this notion, when the *tpiA* gene was deleted, which diverted carbon flux from DHAP to 1,2-PD, we noticed much improved cell growth and a large increase in 1,2-PD titers (Fig. 6). Hence, our future studies will explore whether the assembly (i.e., covalent via fusion or non-covalent via PIDs) of MgsA and an enzyme further upstream, namely FpaA, will effectively "channel" DHAP to 1,2-PD in a manner that mimics *tpiA* deletion.

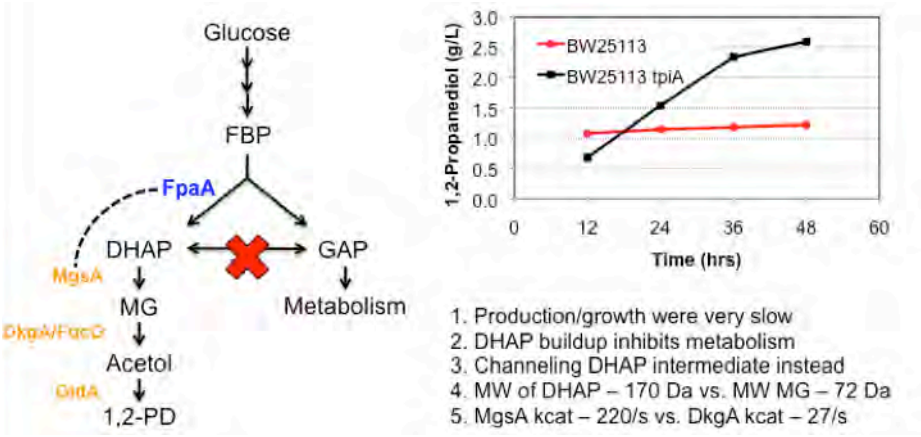

Figure 6. Exploring alternative enzymes to assemble for 1,2-PD production. Deletion of *tpiA* results in 2.5-fold increase in 1,2-PD titers suggesting that assembling FpaA with MgsA via fusion or PIDs might be a viable strategy to increasing 1,2-PD titers.

E3. Employ metabolic sensors for optimizing enzyme assemblies via directed evolution and/or strain engineering. As mentioned above, the improvement in 1,2-PD conferred by the protein scaffold was only about 50% as effective as our best case using the GCN4 PIDs attached to 2 of the pathway enzymes (see Fig. 4). Thus, further optimization of the enzyme-scaffold system is needed. This will be attempted by combining the fluorescent metabolite biosensors, which have already constructed for D-BT (see progress report for grant N00014-07-1-0027) or which are under construction for 1,2-PD, with combinatorial libraries of enzyme-scaffold systems (Fig. 7). Briefly, libraries of the protein scaffold and/or enzyme-ligand constructs will be created by error-prone PCR and co-transformed with the fluorescent metabolite biosensor plasmid. The premise for this approach is that mutations to the scaffold and/or enzyme-ligand constructs that enhance 1,2-PD (or D-BT) titers will lead to an increase in cell-associated fluorescence. These highly fluorescent cells can be easily recovered using a fluorescence-activated cell sorter (housed in the DeLisa laboratory) for library screening. All positive clones will be characterized in detail during the final period of ONR support.

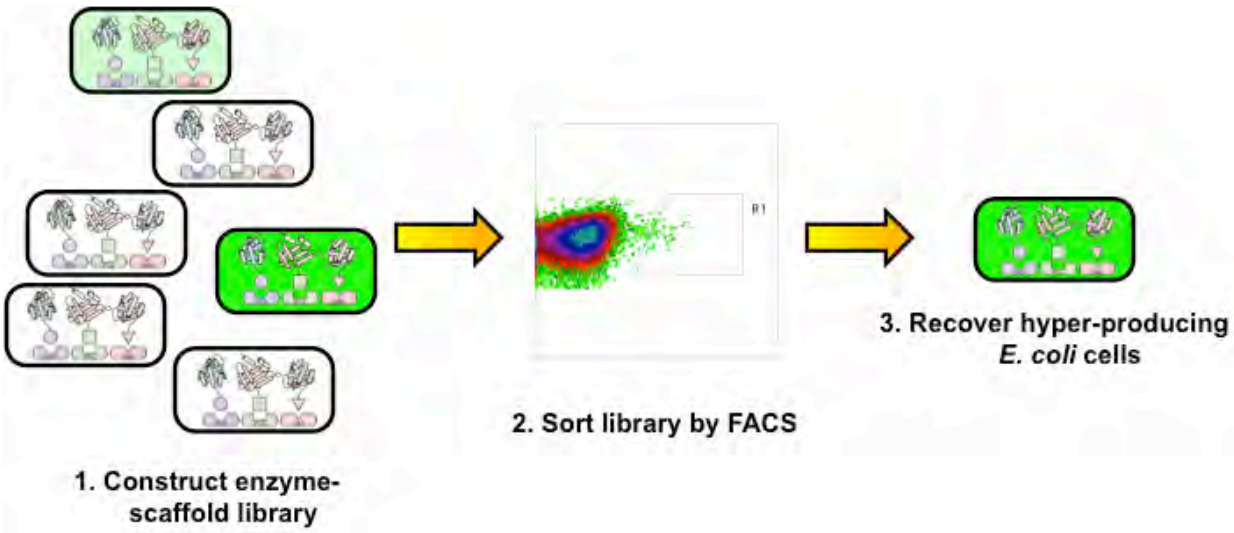

Figure 6. A directed evolution strategy for optimizing enzyme-scaffold systems using fluorescent metabolite biosensors and combinatorial libraries of scaffold and/or enzyme-ligand sequences.

F. Major Problems/Issues. There have been no significant problems.

G. Technology Transfer. We have disclosed one new invention during this period.

Conrado, R.J. and DeLisa, M.P. “Compositions and methods for intracellular enzyme assembly and uses thereof” provisional patent filed.

References Cited

Altaras, N. E. & D. C. Cameron, (1999) Metabolic engineering of a 1,2-propanediol pathway in Escherichia coli. *Appl Environ Microbiol* **65**: 1180-1185.

DeLisa, M. P. & R. J. Conrado, (2009) Synthetic metabolic pipelines. *Nat Biotechnol* **27**: 728-729.

Dragan, A. I., L. Frank, Y. Y. Liu, E. N. Makeyeva, C. Crane-Robinson & P. L. Privalov, (2004) Thermodynamic signature of GCN4-bZIP binding to DNA indicates the role of water in discriminating between the AP-1 and ATF/CREB sites. *Journal of Molecular Biology* **343**: 865-878.

Dueber, J. E., E. A. Mirsky & W. A. Lim, (2007) Engineering synthetic signaling proteins with ultrasensitive input/output control. *Nat Biotechnol* **25**: 660-662.
Dueber, J. E., G. C. Wu, G. R. Malmirchegini, T. S. Moon, C. J. Petzold, A. V. Ullal, K. L. Prather & J. D. Keasling, (2009) Synthetic protein scaffolds provide modular control over metabolic flux. *Nat Biotechnol* **27**: 753-759.
Heuer, K. H., J. P. Mackay, P. Podzebenko, N. P. S. Bains, A. S. Weiss, G. F. King & S. B. EasterbrookSmith, (1996) Development of a sensitive peptide-based immunoassay: Application to detection of the Jun and Fos oncoproteins. *Biochemistry* **35**: 9069-9075.
Liu, Q., D. J. Segal, J. B. Ghiara & C. F. Barbas, 3rd, (1997) Design of polydactyl zinc-finger proteins for unique addressing within complex genomes. *Proc Natl Acad Sci U S A* **94**: 5525-5530.
Mandell, J. G. & C. F. Barbas, 3rd, (2006) Zinc Finger Tools: custom DNA-binding domains for transcription factors and nucleases. *Nucleic Acids Res* **34**: W516-523.
Oyama, R., H. Takashima, M. Yonezawa, N. Doi, E. Miyamoto-Sato, M. Kinjo & H. Yanagawa, (2006) Protein-protein interaction analysis by C-terminally specific fluorescence labeling and fluorescence cross-correlation spectroscopy. *Nucleic Acids Research* **34**: -.
Patel, L. R., T. Curran & T. K. Kerppola, (1994) Energy-Transfer Analysis of Fos-Jun Dimerization and DNA-Binding. *Proceedings of the National Academy of Sciences of the United States of America* **91**: 7360-7364.
Pernelle, C., F. F. Clerc, C. Dureuil, L. Bracco & B. Tocque, (1993) An Efficient Screening Assay for the Rapid and Precise Determination of Affinities between Leucine-Zipper Domains. *Biochemistry* **32**: 11682-11687.
Sceller, L. L., C. Ripoll, M. Demarty, A. Cabin-Flamand, T. Nystrom, M. S. Jnr. & V. Norris, (2000) Modelling bacterial hyperstructures with cellular automata. *Interjournal Paper 366*.
Segal, D. J., B. Dreier, R. R. Beerli & C. F. Barbas, 3rd, (1999) Toward controlling gene expression at will: selection and design of zinc finger domains recognizing each of the 5'-GNN-3' DNA target sequences. *Proc Natl Acad Sci U S A* **96**: 2758-2763.